PERSONALITY AND PERFORMANCE

Jeffrey Wiley Crews

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

PERSONALITY AND PERFORMANCE

by

Jeffrey Wiley Crews

and

Jeffrey William Cummings

June 1974

Thesis Advisor: D. E. Neil

Approved for public release; distribution unlimited.

T 161529

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Personality and Performance		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis June 1974
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Jeffrey Wiley Crews Jeffrey William Cummings		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE June 1974
		13. NUMBER OF PAGES 65
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

16PF   GRAPES
Personality Measures
Performance Measures
Naval Academy Graduate Performance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Sixteen Personality Factors Questionnaire (16PF) and the Performance Evaluation Report (PER), two survey devices used by the U. S. Naval Academy were critically evaluated. The hypothesis that personality as measured by the 16PF could be used to predict future fleet performance as measured by the PER was investigated. This analysis suggested little that would indicate a relationship between personality and performance as they are measured by the two surveys. Performance and personality could be

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
(Page 1)   S/N 0102-014-6601 |

related, but the important aspects of these qualities perhaps are not being measured by the survey devices. The motive for seeking a predictive relationship was also addressed. Within the context of the Academy's current training program, the discovery of such a predictive relationship would have little benefit in assessing how well the Academy is preparing officers for fleet duties. A job descriptive inventory of junior officer duties, and evaluation of graduates in these areas would better satisfy that purpose.

Personality and Performance

by

Jeffrey Wiley Crews
Ensign, United States Navy
B.S., United States Naval Academy, 1973

and

Jeffrey William Cummings
Ensign, United States Navy
B.S., United States Naval Academy, 1973

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1974

ABSTRACT

The Sixteen Personality Factors Questionnaire (16PF) and the
Performance Evaluation Report (PER), two survey devices used by the U. S.
Naval Academy were critically evaluated. The hypothesis that personality
as measured by the 16PF could be used to predict future fleet performance
as measured by the PER was investigated. This analysis suggested little
that would indicate a relationship between personality and performance
as they are measured by the two surveys. Performance and personality
could be related, but the important aspects of these qualities perhaps
are not being measured by the survey devices. The motive for seeking
a predictive relationship was also addressed. Within the context of the
Academy's current training program, the discovery of such a predictive
relationship would have little benefit in assessing how well the Academy
is preparing officers for fleet duties. A job descriptive inventory of
junior officer duties, and evaluation of graduates in these areas would
better satisfy that purpose.

TABLE OF CONTENTS

I. INTRODUCTION

The Graduate Performance Evaluation System (GRAPES) was designed for the expressed purpose of evaluating the performance of graduates from the U. S. Naval Academy. The primary measuring instrument used is the Performance Evaluation Report (PER) which is a nonprojective, closed-response type questionnaire. It has been previously analyzed in terms of the most available criteria; the aptitude and academic averages established at the Academy. Unfortunately, no strong predictor of performance has been found.

The Sixteen Personality Factor Questionnaire (16PF), a psychological test administered to all midshipmen upon entrance to the U. S. Naval Academy, is another possible predictor of performance. On an intuitive level the hypothesis seems plausible that an individual's performance could be reflected by his personality profile, provided each area is properly measured. Therefore, the relevant question is, can the 16PF be used to predict a graduate's performance? Or alternatively, is there a desirable personality profile which will enable the anticipation of and solution to problems prior to graduation?

It is the intent of this study to investigate the hypothesis that the 16PF can be used as a predictor of performance as reflected by the PER. First, the 16PF will be described and its psychometric properties discussed. Next, the PER will be described and analyzed with respect to questionnaire design criteria. By paying particular attention to the criticisms and assumptions discussed in these two sections, the results of the last section, "The 16PF as a Predictor of Performance," seems reasonable.

II. THE SIXTEEN PERSONALITY FACTOR QUESTIONNAIRE

A. DESCRIPTION OF THE 16PF

The Sixteen Personality Factor Questionnaire (16PF) is designed to provide information about an individual's personality profile. Its scales are carefully oriented to basic concepts in human personality structure, keeping in mind the "personality sphere concept." In other words, according to Raymond B. Cattell, the creator of the 16PF, a comprehensive coverage across all dimensions of personality is attempted. The fact that twenty-three (sixteen primary and seven secondary) out of a possible thirty are actually measured would seem to indicate a fairly thorough accomplishment of this objective.

Diversity within the field of personality development has created a certain amount of confusion in regards to terminology. The 16PF has attempted to counter this problem by supplementing a technical description of each factor with a universal index symbol and a more common label. This attempts not only to alleviate the problem within the psychological field itself, but also allows for improved communication between psychologists and the lay public.

An understanding of the composition of a factor scale and its corresponding value is necessary. Basically, each scale is comprised of a set of items which correlates significantly with that factor, though not necessarily between items. In this context an item refers to a particular question on the questionnaire: e.g.

```
Do you tend to get angry with people                      IN
        rather easily?                          YES    BETWEEN    NO
```

After utilizing correlational techniques to assign all items of the questionnaire to their respective factors, the next step is to assign to each factor its appropriate score as reflected by the questionnaire results. Unweighted raw scores are easily computed by assigning a zero, one, or two to each item, depending on the response. Then, with some loss of information, a standardization process called sten (standard ten) is imposed. Actually, this process entails two steps. First, a standard-sten is used where the raw score mean of the population is assigned the central value of 5.5. From this point, the scale increments one sten for each half standard deviation of raw score (FIGURE 1).

THE STEN RANGE

$-2\ 1/4\sigma$			$-3/4\sigma$	$-1/4\sigma$	$1/4\sigma$	$3/4\sigma$			$2\ 1/4\sigma$
1	2	3	4	5	6	7	8	9	10
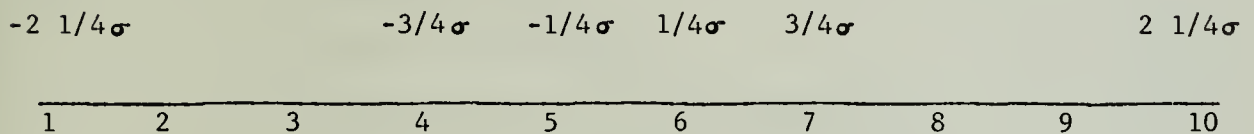
FIGURE 1

Since raw scores tend to yield skewed distributions, a second step is necessary. Through application of a normal transformation the standard-sten becomes a normal-sten, thereby eliminating any skewness while insuring smoothness across the entire range of one to ten. Of course, such transformation guarantees a normally distributed population of scores and equal intervals on which to measure them. Therefore, parametric statistical procedures are applicable in attempting any type of diagnostic or predictive procedures.

The following charts are included for the purpose of associating each factor with its technical psychological title and its more common label (5).

8

TABLE I

TECHNICAL DESCRIPTION OF EACH FACTOR

	Factor	Low Sten Score (1 to 3) vs High Sten Score (8 to 10)
PRIMARY FACTORS	A	Sizothymia vs Affectothymia
	B	Low Intelligence vs High Intelligence
	C	Ego Weakness vs Higher Ego Strength
	E	Submissiveness vs Dominance or Ascendance
	F	Desurgency vs Surgency
	G	Low Superego Strength vs Superego Strength
	H	Threctia vs Parmia
	I	Harria vs Premsia
	L	Alaxia vs Protension
	M	Praxernia vs Autia
	N	Naivete vs Shrewdness
	O	Untroubled Adequacy vs Guilt Proneness
	Q_1	Conservatism of Temperment vs Radicalism
	Q_2	Group Dependency vs Self-Sufficiency
	Q_3	Low Self-Sentiment Integration vs High Strength of Self-Sentiment
	Q_4	Low Ergic Tension vs High Ergic Tension
SECONDARY FACTORS	Q_I	Invia vs Exvia
	Q_{II}	Adjustment vs Anxiety
	Q_{III}	Pathemia vs Cortertia
	Q_{IV}	Subduedness vs Independence
	Q_V	Naturalness vs Discreetness
	Q_{VI}	Cool Realism vs Prodigal Subjectivity
	Q_{VII}	Low Intelligence vs High Intelligence

9

TABLE II

LESS TECHNICAL DESCRIPTION OF EACH FACTOR

	Factor	Low Sten Score (1 to 3) vs High Sten Score (8 to 10)
PRIMARY FACTORS	A	Reserved vs Outgoing
	B	Less Intelligent vs More Intelligent
	C	Affected By Feelings vs Emotionally Stable
	E	Humble vs Assertive
	F	Sober vs Happy-Go-Lucky
	G	Expedient vs Conscientious
	H	Shy vs Venturesome
	I	Tough-Minded vs Tender-Minded
	L	Trusting vs Suspicious
	M	Practical vs Imaginative
	N	Forthright vs Shrewd
	O	Placid vs Apprehensive
	Q_1	Conservative vs Experimenting
	Q_2	Group-Dependent vs Self-Sufficient
	Q_3	Undisciplined Self-Conflict vs Controlled
	Q_4	Relaxed vs Tense
SECONDARY FACTORS	Q_I	Introversion vs Extraversion
	Q_{II}	Low Anxiety vs High Anxiety
	Q_{III}	Responsive Emotionality vs Alert Poise
	Q_{IV}	Dependence vs Independence
	Q_V	Less Neurotic Trend vs More Neurotic Trend
	Q_{VI}	Less Leadership Potential vs More Leadership Potential
	Q_{VII}	Less Creative Personality vs Creative Personality

In the interest of maintaining a less technical level, subsequent discussions

will refer to the more common labels. A more complete description of each

factor is included in Appendix A. The order of factor presentation,

according to Cattell, is based on evidence of diminishing contribution

to behavioral variance.

A few more points are worth mentioning about the factors and associated

scale positions. First, note that extreme scores, high or low, may not

always be desirable. Statements such as, "low scores are always bad" can

be totally inappropriate. Second, it appears at first glance that some

factors may have been excluded. There are two: Factor D (Phlegmatic

Temperament vs Excitability) and Factor J (Zeppia vs Coasthenia). These

two factors are covered in the HSPQ (High School Personality Questionnaire)

but, according to Cattell, are not vital enough to be displayed by the

16PF for adults.

The secondary factors, as their name implies, serve only secondary

functions, and are not as precisely defined as are the primary factors.

Therefore, a detailed discussion on the level of that associated with the

sixteen primaries is impossible. However, their general purpose and

relationship to the sixteen primary factors will be stated. They serve

as broad influences or organizers contributing to the primaries and account

for any inter-factor correlations which might exist.

B. DESIGN, CONSTRUCTION, AND PSYCHOMETRIC PROPERTIES OF THE 16PF

Any discussion concerning a particular questionnaire or test would be incomplete without mentioning some of the principles incorporated into its design and construction and psychometric properties of the scales themselves.

It is of considerable importance in the use of the 16PF (as with all questionnaires) to insure that good cooperation can be achieved, that distortion and sabotage can be detected, and that the scales selected are appropriate for the educational level of the group to be tested. Fortunately, the last requirement is easily satisfied due to the existence of three sets of parallel forms. Describing their construction briefly, Form A is designed equivalent to B, C to D, and E to F. Forms A and B each have 187 items, requiring 45 to 55 minutes per form for an average reader. They are written at about a seventh-grade reading level, though they are also suitable for college students. In order to insure participation across all factions of society, Forms C and D (fifth-grade level), requiring 20 to 30 minutes to complete, and Forms E and F (third-grade level), requiring 20 to 30 minutes to complete, are available. Equivalent forms (pairs) were designed to allow for testing and retesting of the same individual after a short time period. Three sets were provided so that different socio-educational backgrounds could be compared and so that time would be no factor.

The second point is more difficult to counter because either deliberate sabotage (willfully responding incorrectly to questions) or unconscious motivational role distortion (responding to questions as one believes he is expected to respond) comes into play. Fortunately, statistical techniques which are compatable with the 16PF exist to detect and offset these effects (5).

The first point, and perhaps most important, is the most difficult to insure. Good cooperation depends upon the environment in which the test is administered, and upon the rapport between the subjects and the administrator. Therefore, the responsibility of insuring subject cooperation falls largely on the test administrator.

Other problems must be overcome if validity of results is to be achieved. There is a tendency for response set effects to occur when questionnaires are being answered. In this particular questionnaire these effects are investigated in relation to (1) acquiescence, (2) extremity of response, and (3) social desirability of response. By equalizing the number of items for which "yes" and "no" answers contribute positively to the score on each factor, the first problem is eliminated. The various forms (A, B, C, D, E, and F) can be utilized to insure the existence of extreme responses. Generally, it can be said that the more adequately educated and disciplined a subject is, the more latitude he can be given. Using this reasoning, the correct form can be selected. Consistent with this, Forms E and F follow a forced-choice format (no middle category) where as the other four have all three choices. But the problem of social desirability is dealt with quite differently. It is included in the determination of factor Q_{II}. Therefore, it seems that the developers of the 16PF have made a conscious effort to control response set effects.

One last area is of prime importance in a consideration of the 16PF: the psychometric properties of the scales. By addressing the concepts of reliability and validity, it will become apparent that problems may exist concerning statistical inferences which can be made.

Reliability concerns the agreement of two different administrations of the same test. The construction of the test itself, its mode of

13

administration, and its manner of scoring all contribute in some way to this concept. Conspect reliability (agreement between two scorers) is of no interest here since the test is objectively scored. However, dependability and stability do play a significant role. The former, represented by a dependability coefficient (Table III), is concerned with the correlation between two administrations of the same test within a period of time, insufficient for anyone to change with respect to what is being measured. The latter, represented by a stability coefficient (Table III), is concerned with the same correlation, but after a two-month or longer interval.

It can now be seen that statistical problems might be encountered when projecting the results over a five year interval. The 16PF is administered to midshipmen five years prior to completion of the PER. A look at the stability coefficients indicates that one's personality profile is very receptive to change over such a long time span. Therefore, a very significant simplifying assumption will have to be made (referred to later as "Black Box Assumption") in order to lend any support to any conclusions which might be met.

Transferability, the agreement of what is measured across different populations; validity, the agreement of what is measured with what should be measured, are as important if not more so than reliability. But according to Cattell and some critiques written on the 16PF, the construct and concrete validities are as high, if not higher than any other method for measuring personality, and the test is transferable across a wide variety of populations.

Much criticism has been aimed at the 16PF from various experts in the field of psychology. One common complaint doubts the "claim" that the

TABLE III

RELIABILITY COEFFICIENTS FOR EACH FACTOR
(100 = PERFECT AGREEMENT BETWEEN SCORES)

	DEPENDABILITY COEFFICIENT		STABILITY COEFFICIENT	
	FORM A	FORM B	FORM A ($2\frac{1}{2}$ mo. interval)	FORM A (4 yr. interval)
A	81	75	80	49
B	58	54	43	28
C	78	74	66	45
E	80	80	65	47
F	79	81	74	48
G	81	77	49	54
H	83	89	80	49
I	77	79	85	63
L	75	77	75	40
M	70	70	67	43
N	61	60	35	39
O	79	81	70	57
Q_1	73	70	50	52
Q_2	73	75	57	46
Q_3	62	62	36	41
Q_4	81	87	66	56

items represent an even sampling from the personality sphere with a minimum of overlapping of factor scores. Another concerns the arrangement of the factors. Why can the traits not be arranged in three groups: traits largely determined by heredity, traits largely dependent on environment, and traits related to ego formation? But in the interest of simplicity and convenience, the 16PF will be considered an adequate measure of human personality.

C. TEST ADMINISTRATION

The 16PF was administered to all entrants to the U. S. Naval Academy one week after their arrival. Either Form A or B was utilized. Care was taken to assure that the questionnaire was given in a relaxed environment to enhance the cooperative spirit of the midshipmen. The data available for the analysis consists of 295 personality profiles of 1972 graduates from the U. S. Naval Academy. Fifty profiles were selected at random from this population and an average scale position along with its associated standard deviation was computed for each primary and secondary factor. This profile can be seen in the following table.

TABLE IV

PERSONALITY PROFILE OF "TYPICAL"MIDSHIPMAN

. . .

	MEAN	STAND. DEV.
A	5.08	1.95
B	8.38	1.19
C	5.26	1.94
E	7.22	1.74
F	7.27	2.20
G	4.44	2.10
H	5.89	2.26
I	5.85	2.45
L	6.27	1.84
M	7.03	1.91
N	2.96	1.61
O	5.68	2.42
Q_1	4.89	1.89
Q_2	4.67	2.17
Q_3	4.97	2.21
Q_4	6.58	2.61
Q_I	6.87	2.13
Q_{II}	6.22	2.50
Q_{III}	5.36	1.96
Q_{IV}	6.51	1.92
Q_V	4.97	2.41
Q_{VI}	5.39	2.24
Q_{VII}	6.78	1.94

III. THE PERFORMANCE EVALUATION REPORT

A. DESCRIPTION

The primary instrument used in evaluating performance by the GRAPES program is the Performance Evaluation Report (PER). This report is a questionnaire addressed to the commanding officers of Naval Academy graduates with initial surface line assignments. The commanding officers are asked to rate the graduates after one year of observation in 37 performance catagories and 15 personal characteristics categories. Additionally, the graduate is compared to officers from other sources for performance, professional knowledge, and officer-like qualities within the areas of engineering, operations, deck, and weapons as well as overall performance.

Different rating scales are used for each section of the questionnaire. Within the performance section, the scale ranges from "strong" to "unsatisfactory" with intermediate values of "adequate" and "weak" plus an additional column for "not observed." In the personal characteristics section the scale is arranged so the graduates can be placed into percentage groups with regard to the specific charactertistic. The percentage groups are: top ten percent, next forty per cent, next forty per cent, and bottom ten per cent. A "not observed" column is also included. Within the comparison section the scale ranges from "much better" to "generally worse" with intermediate values of "generally better" and "no significant difference." Again, a "not oberved" column is included.

The categories in which the graduates are to be rated within the performance section of the PER are grouped into five major areas: general, operations, navigation, engineering, and weapons. This division corresponds with the various designations of the officer's primary duty indicated within the heading of the questionnaire. Other information included in this heading is: name of the person to be evaluated, his social security number, name of his command, date of the report, the basis of observation, and general instructions on completing the PER.

The items comprising the performance and personal characteristics sections of the PER are included in Appendix B.

B. QUESTIONNAIRE DESIGN AND THE PER

Abraham Oppenheim (14) in his book Questionnaire Design and Attitude Measurement states that the primary function of a questionnaire is the measurement of a specific set of variables. Performance, the attribute which the PER was designed to evaluate, is a most difficult and elusive quantity to specify with a set of observable variables. The situations and environments into which the graduates are placed and their evaluators are so varied that no widely accepted norms of "performance" exist. In general, there seem to be no familiar and consistent scales on which to measure "performance." Perhaps an inventory and assessment of the jobs for which graduates are responsible during their first year of fleet duty could be conducted. Then the important variables that should be measured by a questionnaire would be identified, and the PER could be designed to reflect the variables.

According to Professor Richard Elster of the Naval Postgraduate School, the United States Coast Guard is currently conducting a job descriptive

inventory for recent graduates of the Coast Guard Academy with the objective of adjusting the curriculum of that institution to emphasize the areas highlighted by the job inventory. Enlisted rates within the Navy are also receiving the same scrutiny through the Navy Occupational Task Analysis Program. However, the construction of the PER does not seem to be based on any such analysis. This was suggested by a small experiment conducted at the Naval Postgraduate School. A list of the areas of evaluation, exactly as they appear on the PER, was distributed to naval officers with experience ranging from division officer to department head. The officers were asked to designate which items they considered to be important in the evaluation of first year fleet performance. There was no significant agreement among the 18 responses returned to the experimenters. One officer, a former chief engineer aboard a destroyer said, "I firmly believe most of these questions concern what an ensign should learn after commissioning. All the Academy should do is give a basis to build on." Another officer commented that "a general knowledge of all these areas would be nice."

Pilot work is another important step in formulating an acceptable questionnaire. Before a questionnaire can be used to gather data, it should first be tested to certify that it is measuring the variables specified within its stated purpose. This testing process identifies such inadequacies as ambiguous questions, poor rating scales, unclear instructions, and inadequate letters of introduction. There is evidence that suggests the PER was subjected to little or no pilot work. One potential indicator of inadequate piloting can be seen by examining the number of "not observed" responses for each item of the PER. Any item with a significant number of "not observed" responses might prove to be

irrelevant, and perhaps should not be included within the questionnaire. Thirteen of the thirty-seven items within the performance section of the PER had a "not observed" response rate of more than one third. In fact, more than two-thirds of the responses for one item fell into the "not observed" column. A table of the "not observed" responses for each item is included in Appendix C.

Another possible inconsistency in the PER that might have been discovered through pilot work can be disclosed by investigating the rating scale used within the personal characteristics section. Recall that in this section of the questionnaire graduates were to be placed within designated percentage groups. However, the distribution of the responses did not at all coincide with the indicated percentage groups of the scale. The 295 PERs of graduates of the class of 1972 disclose that more than 55 per cent of the responses in the personal characteristics section were in the "top ten per cent" scale position while fewer than 44 per cent of the responses were in the middle 80 per cent scale positions and only 1.1 per cent of the responses were in the "bottom ten per cent" scale position. A histogram of the actual response frequencies by scale position is contained in Appendix D.

Oppenheim further states that a questionnaire must be designed to be amenable to specific pre-selected statistical techniques. This means that special care must be taken in designing rating scales. Most parametric statistical measures can only be applied to interval data; while the trouble with most rating scales is that the intervals between various points on the scale are not of equal size. This results in an ordering on the scale rather than exact positioning. The rating scales for both the performance and the personal characteristics sections of the PER

21

appear to have intervals of unequal size. Examination of the histogram of response frequencies for the performance section shows that the two highest points on the scale accounted for more than 91 per cent of the responses; the adequate position accounted for 53 per cent of the responses while the strong position accounted for another 39 per cent. This may indicate that the difference between adjacent points on the scale is not equal; there being a wider gulf between the weak and adequate positions than exists between the adequate and strong positions. A similar discussion has already been presented for the personal characteristics section. Because of the unequal intervals within both scales, the assignment of equally-spaced numerical scores to the different scale positions and the computation of such statistics as means and standard deviations is virtually meaningless. A pilot study would have revealed this fact.

It is most important that the effort to gather data for any study must be designed with utmost care to insure the success of the undertaking. The essential steps of this design process according to Oppenheim are:

1. Decide the aims of the study and the hypotheses to be investigated.

2. Review the relevant literature; discuss with informants and interested bodies.

3. Design the study and make the hypotheses specific to a situation (make the hypotheses operational).

4. Design or adapt the necessary research methods and techniques (the questionnaire in this case); pilot work and revision of the questionnaire.

5. The sampling process: selection of the people to be approached.

6. The field-work stage: data-collection and returns via circulation of the questionnaire.

7. Process the data, code the responses.

8. The statistical analysis; test for statistical significance.

9. Assemble the results and test the hypotheses.

10. Write up the results: relate the findings to other research; draw conclusions and interpretations.

There are other important aspects of questionnaire design. For instance, the "halo effect" must be guarded against. It can occur when all the favorable responses lie in the same column and similarly all the unfavorable responses lie in the same column. This allows the grader to let his general impression of the person he is rating determine which column receives the predominant number of responses. Therefore, the person is not evaluated on each individual item of the questionnaire. In the PER there is some doubt as to whether the "halo effect" was considered since all of the most favorable responses were in the extreme right column, and all of the least favorable responses were in the left column just inside the column for "not observed" responses. One procedure for guarding against this effect would have been to word the items of the survey so that the column of the most desirable response shifts from right to left necessitating the reading of each item to at least identify the location of the favorable (or unfavorable) response. This might have stimulated responses based on the individual's merit for each item.

Another problem generated by the use of rating scales in a questionnaire is to certify that all of the raters have similar perceptions about the qualities to be rated so that they can view them from the same frame of reference. Many of the individual items appearing on the PER might be subject to such perceptual difficulties. For instance, it is not at all guaranteed that attitude, one of the items to be rated in the personal characteristics section, would be viewed the same by any two commanding officers. Similarily, there is no assurance that two commanding officers would agree on what comprises adequate knowledge of the causes and effects of weather, especially if one happens to be a meteorologist while the other is not.

23

Still another aspect of questionnaire design that should be considered in connection with the PER is the form of the response. There are in general, two types of questions: open or free response types, and closed or fixed alternative types. Both have their unique advantages and disadvantages. All of the items on the PER are of the closed response type, with the location on the rating scale representing the fixed alternatives. Some of the advantages of closed response questionnaires over open response types include easier completion and quantification of results, less writing requirements, and the capacity for gathering information in less time for a smaller sum of money. The prime disadvantage of the closed response questions is that closed responses lose much of the thought put into the question by the respondent because he is forced to choose between fixed alternatives. This forced choice might lead to a loss of rapport between the testing agent and the respondent if the respondent feels that none of the alternatives adequately reflects his ideas in that area. In the case where rating scales exist the respondent may even resort to marking column dividing lines, indicating that there should be another choice between two adjacent categories. For instance, although a person's performance on one of the items of the PER might not be "strong" there may be a hesitancy on the part of the commanding officer to mark him as "adequate" if the commanding officer connotes adequate with barely satisfactory and strong with not exceeded. Pilot work can often guard against this problem by first testing the question as an open question. Then provided the responses fall into a small number of categories, the question can be reworded as a closed response type. Otherwise, the question is best left open (14).

One of the major difficulties with the free response type of questionnaire is quantification of the responses. One way such quantification is

accomplished is through a method known as coding. This coding is effected
by an impartial member of the study group. His job consists of classifying
the responses into categories and placing the categories of responses on
a rating continuum. During the coding process much the same information
loss occurs as through closed response questioning. However, since all
of the coding is done by a single individual, problems of differing
perception may be minimized. To be sure, the coder might be biased, but
the bias should be more consistent and more easily identified than the
biases resulting from a nonstandardly perceived rating scale in a closed
response environment. Additionally, through the use of free response
type questions, problems with the perception of the questions might be
uncovered. Some of the prejudices and predispositions of the respondent
that would affect his ratings might appear within the text of his responses.

C. STATISTICAL ANALYSIS OF THE PER

It seems reasonable that efforts should be made to insure effective
utilization of the respondent's time and space on the PER. This might
be accomplished by analyzing the information the PER items yield and
seeing if any of these items, or entire groups of questions, are redundant
in the information they provide. If this should be the case, then the
redundant groups could be eliminated, giving the respondent fewer items
to rank, with more thought devoted to each item. Alternatively, a free
response section could be added to the questionnaire to provide some more
detailed aspects of performance data.

The first step in studying the data obtained from the PER was to
quantify the responses on the rating scales. Ideally, the interval
distance between adjacent points on the scales should be of equal size

25

allowing for the use of interval based as well as ordinal statistics.

One way of artificially producing intervals of equal size is to allow

the empirical distribution of responses to determine what numerical values

to associate with each response category. This empirical cumulative

distribution scaling technique was utilized to evaluate both the performance

and the personal characteristics sections of the PER. The technique was

applied as follows. First, a numerical scale ranging from 0.0 to 4.0 was

selected to be paired with the responses. Then, a cumulative frequency

distribution of responses was formed from the population of 295 PERs. The

distribution began with the least favorable response and compiled succes-

sively toward the most favorable response. Using the empirical cumulative

frequency distribution, the most favorable response was assigned a numerical

value of 1.0 times the maximum scale value 4.0. The next most favorable

response was assigned the value of the cumulative frequency distribution

at that point times the maximum scale rating, and so on. The histograms

for the distributions of responses for the performance and personal

characteristics sections of the PER can be seen in Appendix D, along with

the numerical values for each response.

With the responses quantified in a useful manner, some hypotheses

were made and tested about the data obtained from the PER. Viewing the

histograms of the responses to the individual items and the overall

response histograms for the sections of the questionnaire, one can see

that many of them do not resemble the familiar bell shape of the normal

distribution. For this reason, non-parametric statistical techniques not

requiring the assumption of an underlying normal distribution were

utilized. Since some of the non-parametric analytic schemes are not

easily amenable to computer analysis, a random sample of 50 subjects was

26

drawn from the population of 295 reports to facilitate the hand computation of the statistics. The power of these tests with a sample size of 50 is almost identical to the power of the same tests with an infinite sample size (refer to power curves).

In preparation for the statistical analysis to be conducted, four mean scores were calculated for each of the 50 sample subjects. An overall performance mean was calculated over all of the 37 performance items. Additionally, means were calculated for the general area of the performance section and for the area of primary duty. An overall personal characteristics mean was also computed using all fifteen items in that section of the PER. A table of these averages can be seen in Appendix E.

One of the first bits of information that can be obtained from the questionnaire is a measure of consistency between the ratings within the performance section and the personal characteristics section of the PER. This concept stated in hypothesis form is that there is no significant difference between the overall performance averages and the personal characteristics averages. This hypothesis was tested using the Wilcoxon Matched-Pairs Signed-Ranks Test, one of the most powerful alternatives to parametric tests. The results supported the hypothesis that there is indeed no significant difference between the performance averages and the personal characteristics averages. Having determined that the personal characteristics and performance averages yield essentially the same results relative to a performance index, another tack might be to see if certain sub-sections of the performance section, specifically the general and primary duty areas, yield a performance index comparable with the personal characteristics section. This idea stated in hypothesis form is that there is no significant difference among the averages of the general sub-section, the primary duty sub-area, and the personal characteristics section of the PER.

27

This hypothesis was tested using the Friedman two-way analysis of variance. The results of this test supported the hypothesis that there is no significant difference among the averages of the general sub-section, the primary duty sub-area, and the personal characteristics section of the PER.

It might now be of interest to see which performance averages are most highly associated with the personal characteristics section of the PER. To determine this, the Spearman rank correlation coefficient measure of association was calculated for the overall performance--personal characteristics pair. Then Kendall's coefficient of concordance was calculated to measure the degree of association among the general sub-section, the primary duty sub-area, and the personal characteristics section. The coefficient of concordance was then converted to an equivalent value of the rank correlation coefficient for comparison. The results of this test show that the overall performance averages and the personal characteristics averages have a slightly higher degree of association than do the general, primary duty and personal characteristics averages. However, the degree of association is statistically significant in both cases. It therefore appears that as far as calculating a performance index from the data of the questionnaire is concerned, any of these averages is sufficient and comparable to all the others. Complete numerical results of the statistical tests performed on the PER are included in Appendix F.

Because of the great number of items within the PER that had a significant number of not observed responses, use of the overall performance average might not be the best approach. However, of the items included in the general sub-section of performance the highest not observed rate was 12 per cent, with most of the items having not observed rates of around one per cent. Also, it is plausible that the subjects are scrutinized

most carefully in their area of primary duty. With this in mind, a wise decision might be to utilize the averages from either one of these two sub-areas as a performance index. The personal characteristics average probably is not as stable a measure of performance, per se, because the items within that section are more personality than performance oriented.

These results seem to indicate that if an index of performance is the objective of the PER, then it can be considerably simplified to include only those items in the general sub-area. Or the commanding officers can be asked to evaluate the graduates in only their primary duty area. This narrowing of the scope of the PER could also be accomplished by the addition of some free-response questions about the graduates' performance in general. There may be other purposes to be served by the PER. If so, they should be stated explicitly and perhaps assigned as the object of another subsidiary study, for a questionnaire serving too many purposes may end up serving none well.

D. FURTHER REMARKS

Any revision of the PER should be carefully piloted before it is officially used as a data collection instrument. Perhaps this piloting effort could result in modifications to the interval descriptions of the rating scales in order to insure adequate and equally spaced response alternatives. Some of the open-question responses might point to dimensions of performance that have been heretofore overlooked by the items of the PER. Certainly the open-questions would allow some contribution from the experience of the various commanding officers to add to the effectiveness of the study. When sent to the commanding officers for completion, the PER should be accompanied by a letter of introduction

29

explaining the purpose of the study and eliciting his most sincere coopera-
tion. This letter of introduction additionally needs to be piloted before
its actual use in the study to insure that it is fulfilling its intended
purpose.

A questionnaire to gather data should not be assembled without
considerable effort on the part of the group conducting the analysis.
Careful planning must prevail throughout the process beginning with
identifying the exact purpose of the study and the variables to be
measured by the questionnaire, and continuing through the interpretation
of the results of the statistical test completed on the gathered data.
The study must be viewed with a systems approach. All aspects of the
endeavor, especially the ways they interact with one another, must be
considered in the design of the analysis. And the formulation of the
questionnaire is but a single step in analyzing the problem at hand.

IV. THE 16PF AS A PREDICTOR OF PERFORMANCE

A. ASSUMPTIONS

The 16PF and the PER have been previously discussed in great detail.
It has been shown that neither is, by any means, perfect. However, for
purposes of this section, each will be assumed to measure with some
objectivity its respective area. The question at hand here is can the
16PF be used to predict performance?

The 16PF is administered some five years before the results of the
PER are compiled. Since the coefficients of stability are low for all
of the factors of the 16PF, it seems unlikely that scores on a follow-up
administration of the 16PF coinciding with the circulation of the PER
would correlate at all with the scores of the first administration. For
this reason, it must be assumed that the experiences the individuals
encounter during the intervening time between the administration of the
16PF and their subsequent evaluation on the PER are similar with respect
to their effects on personality. Hence, the Academy training program and
environment must be considered equivalent for all individuals. The affect
this program has on the individual is dependent on his personality at the
program's outset as measured by the 16PF.

Analogously, the Academy can be thought of as a black box with inputs
and outputs. These inputs are the people entering the program and the
outputs are the graduates. Assuming that the black box subjects each
input to the same behavior modification process implies the differences in
the output of the system are a function only of the differences in the

31

system's inputs. Hence, the implicit assumption is made that effects of the Academy program can be correlated with the personalities of the incoming midshipmen.

B. STATISTICAL ANALYSIS

The first attempt to uncover a relationship between personality and performance was through utilization of scatter diagrams for each factor of the 16PF, plotting the factor scores against the overall performance averages. Next, each factor score was plotted against the personal characteristics averages. The scatter diagrams indicate that no significant regressional relationship links any of the personality factors individually to performance as measured by the performance section or personal characteristics section of the PER. Multivariant plotting was not attempted because of the perceptual difficulties encountered when more than two dimensions are to be plotted on a plane. Further, multi-variant regressional techniques were not pursued because there were ultimately 23 independent variables which could enter the picture. With a sample size of only 50, no adequate statistical testing could be accomplished.

Having been unsuccessful in determining an overall relationship between personality and performance, a less complicated hypothesis was investigated. Perhaps the 16PF could be used to predict, or at least discriminate between high and low performers. To test this hypothesis, the population of 295 PERs was canvassed, and reports of high and low performers as measured by both the overall performance averages and the personal characteristics averages were extracted for study. The limits for high scores and low scores in each section were arbitrarily selected

with the prime criterion being the sample size. The appropriate cut-off points in the performance section were at 3.70 and 2.00. There were 22 scores above 3.70 and 26 scores below 2.00. In the personal characteristics section there were 49 perfect scores (4.00) and 25 scores below 1.56.

A series of three statistical tests was used to attempt to locate differences in personality factors between high and low performers, determined first by the overall performance averages, and then by the personal characteristics averages. First, a Kolmogorow-Smirnov two sample test (K-S test) for each factor was used to detect any differences in the distributions of the factor scores. It revealed that there were significant differences in factor scores between high performers and low performers as measured by both the overall performance averages and the personal characteristics averages for only a single factor, Factor G: expedient versus conscientious. Since the grouping of data required for the application of the K-S test causes some information to be lost, the Mann-Whitney U test, designed to determine if two samples are drawn from the same population, was applied to the groups of high and low performers. The Mann-Whitney test also indicated that scores for Factor G were not the same for high and low performers as measured by both the overall performance average and the personal characteristics average. Additionally, the Mann-Whitney test indicated that there were also significant differences in the scores of Factor E, humble versus assertive, and Factor Q_{IV}, dependence versus independence between the high and low performers as measured by the personal characteristics average. A parametric t-test was also performed on this data for the following reason: The normalized-sten scoring system imposed on the 16PF factors insures a normal distribution of scores. Although the samples in contention here were not

randomly drawn, there was evidence (discussed previously) indicating that the performance score was not highly correlated with any specific factor score. Therefore, selection of a sample based on performance scores may still have resulted in a random sampling of the population. The t-test yielded the same results as did the Mann-Whitney test with the exception that the t-test did not reveal any difference in the scores on Factor E between high and low performers as measured by the personal characteristics average. Perhaps the controversial randomness assumption causes this apparent loss of power. Thus, it appears that the Mann-Whitney test is the most powerful to use in this situation. A review of the statistical techniques used in this analysis along with all of the numerical results can be seen in Appendix G.

C. DISCUSSION OF RESULTS

The results of this factor by factor analysis seem to indicate that the personal characteristics scores are more influenced by personality than are the overall performance scores. But it seems that neither is influenced drastically enough by differences in personality to permit the 16PF to be used as a predictor or differentiator of performance extremes. Even the consistent significant difference in scores between high and low performers in Factor G has no real predictive value because persons with intermediate performance scores can have scores over the entire rating range for Factor G. So, though it would be nice to be able to say a score of " " on Factor " " means " ", it is impossible considering the method just described.

Considering the factors one at a time does not account for possible patterns of overall personality that could be similar among the different ranges of performance scores. Cluster analysis can be used to detect

34

such patterns in multi-dimensional spaces. However, due to the small
number of elements in some of the samples and the correspondingly large
number of casual factors, cluster analysis is not statistically valid.
Perhaps when more data is collected, and larger samples of performance
groups are accumulated, cluster analysis can be applied to the problem.
Certain numerical techniques do exist that would enable multi-dimensional
clusters or groups to be located. One technique utilizes the projection
of points in multi-dimensional space onto a two-dimensional plane. Through
rotation of the plane of the projection, clusters can be separated. The
method is one of trial and error, and for this reason, it also has little
statistical validity and would not be useful in predictive situations.

Failure of these statistical methods to link performance with person-
ality could indicate that the two are unrelated. On the other hand, this
result could also be the product of several other factors in isolation
or acting together. The 16PF and the PER were not designed specifically
to be used in conjunction with one another. The effects of the normal
transformation of the factor scores on the 16PF could have masked possible
relationships between the factors and performance. If certain aspects of
personality do affect facets of performance, perhaps the PER is not
adequately measuring these particular facets. Whatever the reasons for
the largely negative results of this analysis could be, they cannot be
exactly pinpointed because of the poor design of the data gathering devices.

One must also critically examine the utility of predicting the future
performance of men already admitted to the Naval Academy. After all,
initial screening procedures prevent persons with personalities incompatible
to life within the military environment from being admitted to the Naval
Academy. Therefore, one might assume that those individuals admitted to

35

the Naval Academy possess personalities that would allow them to succeed
in a military environment. If this is indeed the case, then one must
doubt the importance of being able to predict the level of fleet performance
of individuals already admitted. On the other hand, it is recognized
that the need to detect future problems despite accurate screening proce-
dures is ever present.

Suppose the Academy was considering a new program; one which would not
subject all inputs to the same behavior modification. Instead, it would
be tailored for each individual on the basis of his personality. In this
case the ability to predict future performance based on the input person-
ality would be most useful. But, suppose the Academy is interested in how
well its current program is preparing the graduates for their jobs in the
fleet. Here, a prediction of performance based on entering personality is
really not important. Feedback is needed here on the general level of
performance of the Naval Academy graduate. It is in this situation where
the PER information can be most useful, provided the PER is gathering data
on the relevant aspects of first year officer performance.

Currently, it appears as though the PER is designed to measure "how
well are midshipmen learning what the Academy is putting forth." This
is not the relevant question. Instead, the PER should be seeking to
discover "is the Academy teaching the correct areas" and then to probe
into how well things are being presented. Once again, the need for a job
inventory is stressed so that the relevant areas can be identified. Then,
perhaps, the GRAPES program can yield some useful results, rather than a
mass of statistics with dubious implications.

V. SUMMARY

The 16PF and PER were reviewed as measures of personality and performance, respectively. Although there is some controversy concerning whether or not the 16PF accurately measures all aspects of personality, it has been assumed that the test does for the purposes of this analysis. The PER measures performance on 37 items that parallel the U. S. Naval Academy's present curriculum.

There is no apparent relationship between personality and performance as measured by the respective questionnaires. Poor design of the PER combined with inappropriate use of the 16PF seems to be the best explanation. It is recognized that there exists a need to anticipate and remedy any individual's problems before graduation. But it seems to be highly unlikely that the 16PF would reflect such information. After all, extreme scores on many factors imply serious disorders, and screening techniques for gaining admittance to the U. S. Naval Academy are designed to counter any such abnormalities.

Much has been said on the proper design of a questionnaire. It has been implied that the design of the PER possibly violates many of the necessary principles. This might cause serious distortions in the end result. But there can be no more serious distortion than to design a questionnaire which is incompatible with the stated objectives. It is suggested that at this time the people responsible for the promotion of GRAPES should reevaluate and specify their intentions. How well the U. S. Naval Academy is teaching the present curriculum appears irrelevant. The important question is "Are the right courses being taught?" Only then can one concern himself with "How well?"

FACTOR DESCRIPTIONS

The following capsule descriptions of each factor are extracted from a memeograph report supplied by Dr. Montor, a professor at the U. S. Naval Academy.

Factor A: Reserved vs Outgoing

The person who scores low on Factor A tends to be stiff, cool, skeptical, and aloof. He prefers things to people, working alone, and avoiding compromises of viewpoints. He is likely to be precise and "rigid" in his way of doing things and in personal standards; in many occupations these are desirable traits. However, at times he may tend to be critical, obstructive or hard. On the other side of the scale, a high scorer tends to be good natured, easy-going, emotionally expressive, ready to cooperate, attentive to people, and adaptable. He likes occupations dealing with people, thereby rendering him more generous in personal relations. Also, he is less afraid of criticism and more apt to form active groups.

Factor B: Less Intelligent vs More Intelligent

A low score on Factor B indicates a tendency to be slow in learning and grasping, dull, and quite receptive to concrete and literal interpretations. Conversely, a high score reflects a fast learner who is quite able to grasp ideas. Needless to say, one's level of culture and alertness is reflected by this particular factor.

Factor C: Affected by Feelings vs Emotionally Stable

A low score on Factor C is common to almost all forms of neurotic and some psychotic discorders. The low level in frustration tolerance for unsatisfactory conditions, the tendency to evade necessary reality

demands and become easily emotional and annoyed, and the accompanying neurotic symptoms (phobias, sleep disturbances), all point towards this fact. The person who scores high tends to be emotionally mature, stable, realistic about life, unruffled and consequently able to maintain solid group morale.

Factor E: Humble vs Assertive

The person who scores low on Factor E tends to give way to others, to be docile, and to conform. He is often dependent, confessing, and anxious for obsessional correctness. A high score presents a different picture. Assertive, self-assured, independent-minded, austere, hostile, and extra punitive are all descriptions of an individual in this category. Basically, he becomes a law to himself with total disregard for all authority.

Factor F: Sober vs Happy-Go-Lucky

A low score on Factor F indicates a sober, dependable person who tends to be restrained, reticent, and introspective. Sometimes pessimistic and often unduly deliberate, he is usually considered smug and primly correct by observers. Conversely, a high scorer tends to be cheerful, active, talkative, frank, and carefree. He is frequently chosen as an elected leader. However, he may be a bit impulsive at times.

Factor G: Expedient vs Conscientious

A low score on Factor G is indicative of a person who evades rules and feels few obligations. Consequently, he is often casual and lacking in effort for group undertakings and cultural demands. A high score reflects a conscientious and moralistic individual who is dominated by a sense of duty. It is no wonder that he prefers hard-working people to witty companions.

Factor H: Shy vs Venturesome

A "wallflower" has been used to describe a person who scores low on Factor H. He tends to be slow in speech and in expressing himself, dislikes occupations with personal contacts, and is usually quite unaware of all that is going on around him. Though one who scores high is sociable, bold, inventive, and abundant in emotional response, he can be careless of detail, ignore danger signals, and tend to be "pushy."

Factor I: Tough-Minded vs Tender-Minded

Masculine, realistic, practical, independent, and responsible all adequately describe one who scores low on Factor I. However, he is also skeptical of subjective cultural elaborations, unmoved, cynical, hard, and operates on a "no-nonsense" basis. A high scorer though, tends to slow up group performance and upset group morale by unrealistic fussiness. His day-dreaming, fastidious, and feminine manner prove quite destructive.

Factor L: Trusting vs Suspicious

A low score on Factor L refers to a good team worker who tends to be free of jealous tendencies, adaptable, cheerful, and uncompetitive. A high scorer tends to be mistrusting and doubtful, involved in himself and very self-opinionated. One might suspect him to be a poor team member.

Factor M: Practical vs Imaginative

Though unimaginative, a low scorer on this factor is concerned over detail and is able to keep his head in emergencies. Conversely, a high scorer is likely to be rejected in group activities because of his lack of concern over everyday matters and obliviousness to particular people and physical realities.

Factor N: Forthright vs Shrewd

Unsophisticated, sentimental, and simple adequately describe a low scorer on this factor. Though sometimes crude and awkward, he is easily

40

pleased and content with what comes, and is natural and spontaneous. A high scorer, hardheaded and analytical, has an intellectual and unsentimental approach to situations. Polished, experienced, wordly, and shrewd, he has an approach somewhat akin to cynicism.

Factor O: Placid vs Apprehensive

Though resilient and secure in self-assuredness, a low scorer on Factor O tends to be insensitive to alienation from a group. This results in antipathies and distrust. On the other hand, a high scorer tends to be depressed, moody, and full of worry, to the point where he feels unaccepted in group activities.

Factor Q_1: Conservative vs Experimenting

A low scorer tends to oppose and postpone change, is partial to tradition, and is uninterested in intellectual thought. This results in the insistence on "tried and true" methods, even when something else might be better. The high scorer is more well informed, less inclined to moralize, and more tolerant of inconvenience and change. He tends to be interested in intellectual matters and has doubts about fundamental issues.

Factor Q_2: Group-Dependent vs Self-Sufficient

A low scorer on Factor Q_2 is obsessed with the need for social approval and admiration to the point where individual resolution is lacking. Though he may not necessarily be gregarious by choice, he needs group support. A high scorer is obviously accustomed to making decisions and taking action on his own. It is not that the dislikes people, but rather does not need their agreement or support.

Factor Q_3: Undisciplined Self-Conflict vs Controlled

A low scorer on Factor Q_3 is definitely maladjusted for he will not be bothered with will control and regard for social demands. It follows,

then, that he is not overly considerate, careful, or painstaking. On the other hand, a high scorer is inclined to be socially aware and careful, and evidences "self-respect" and regard for social reputation. He sometimes tends, however, to be obstinate.

Factor Q_4: Relaxed vs Tense

Sedate, tranquil, satisfied, and relaxed all adequately describe the low scorer on this factor. Unfortunately, in some cases, laziness and low performance may result as low motivation produces little trial and error. Conversely, a high scorer tends to be tense, excitable and restless, which ultimately leads to frustration in group encounters.

GENERAL

JUNIOR OFFICER DUTIES (knowledge of division officer and other junior officer administrative duties)

WATCH DUTIES (understanding of watch officer responsibilities and ability to carry them out)

SHIPBOARD NOMENCLATURE (ability to identify and describe components of the ship's structure and major fittings)

SHIPBOARD ORGANIZATION (knowledge of ship, department and division administrative organization, battle organization and watch organization)

NAVAL ORGANIZATION (knowledge of operational and administrative chains of command and functions of each)

MATERIAL MANAGEMENT (knowledge of the 3M system and ability to apply basic management techniques to utilize effectively time and material)

SUPPLY (ability to effectively use the naval supply system)

MILITARY JUSTICE (basic knowledge of military judicial system including JAG manual investigations)

OPERATIONS

CIC OPERATION (knowledge of CIC team, CIC equipment, CIC procedures)

CICWO DUTIES (knowledge of CIC watch organization, CIC publications and CIC watch procedures)

MANEUVERING BOARD (ability to apply maneuvering board techniques correctly and rapidly)

AAW WEAPON SYSTEMS (knowledge of basic AAW weapons team, equipment and procedures)

RADAR SYSTEMS (knowledge of the basic principles of operation of search and fire control radars)

RADIO SYSTEMS (knowledge of basic principles of operation of electronic communications equipment)

METEOROLOGY (knowledge of causes and effect of weather)

RADIOTELEPHONE PROCEDURES (ability to conduct effective, proper voice communications)

SEARCH TECHNIQUES (knowledge of basic search and detection theory and its application)

SECURITY (knowledge of classification, stowage, and handling of classified information and material)

TACTICS (knowledge of and ability to use ATP 1A, Vol. I and II)

NAVIGATION

CELESTIAL NAVIGATION (ability to use tools and publications to navigate by celestial means)

ELECTRONIC NAVIGATION (familiarity with and ability to utilize effectively, information from current electronic aids to navigation)

TERRESTIAL NAVIGATION (ability to navigate by dead reckoning or piloting)

RULES OF THE ROAD (ability to apply the nautical rules of the road in all situations)

SHIPHANDLING (knowledge of standard commands and ability to conn a ship alongside another ship or while mooring and unmooring)

ENGINEERING

SHIP PROPULSION SYSTEMS (knowledge of basic principles and operation of power generation in main shipboard power plants)

AUXILIARY MACHINERY (knowledge of basic operating and maintenance principles of refrigeration, and other auxiliary systems)

DAMAGE CONTROL (knowledge and understanding of basic damage control concepts)

ELECTRICITY (knowledge of A.C. and D.C. circuits, measurements, definitions of terms, knowledge of generating and distribution systems)

IC SYSTEMS (knowledge of sound powered phone procedure, IC systems operation and maintenance)

ENGWO DUTIES (knowledge of engineering watch organization and duties of the engineer watch officer)

DCA DUTIES (knowledge of damage control organization and duties of the DCA)

ASW WEAPON SYSTEMS (knowledge of basic ASW weapons team, equipment, and procedures)

GUN SYSTEMS (knowledge of principles of operation of gun systems and ammunition)

WEAPONS

MISSILE SYSTEMS (knowledge of missile control system, missile guidance, and missile warheads)

SONAR SYSTEMS (knowledge of the principles of operation of SONAR equipment)

FIRE CONTROL (understanding of fire control problem and operation of associated equipment)

SEAMANSHIP (knowlege of shipboard evolutions, such as replenishment at sea, mooring, boat etiquette)

PERSONAL CHARACTERISTICS

ATTITUDE (a positive state of mind toward his command and the Naval Service manifested by interest, motivation, and cooperation)

BEARING AND DRESS (correctness of uniform, smartness of appearance expected of an officer and gentleman)

GROWTH POTENTIAL (capacity to handle jobs of increasing scope and responsibility, the ability to learn and profit from experience)

INDUSTRY (zeal exhibited and energy applied in the performance of his duties)

LOYALTY (his faithfulness and allegiance to his superiors, the service, and the nation)

MATURITY (ability to develop correct and logical conclusions and to act rationally and decisively within the limits of his assigned authority)

MORAL COURAGE (to do what he ought to regardless of the consequences)

PERSONAL BEHAVIOR (his demeanor, disposition, sociability, sobriety and personal habits)

PERSONNEL MANAGEMENT (LEADERSHIP) (faculty of controlling and influencing others in definite lines of direction and maintaining discipline)

PHYSICAL FITNESS (physical stamina, alertness and endurance)

READING ABILITY (reading comprehension, ability to understand material by reading it)

RELIABILITY (can be depended upon to meet his responsibilities and is punctual)

SELF-ASSURANCE (self-reliance, self-confidence, boldness of action)

SELF-EXPRESSION (ORAL) (ability to express himself orally)

SELF-EXPRESSION (WRITTEN) (ability to express himself in written communications, reports, etc.)

APPENDIX C

FREQUENCY OF RESPONSES FOR EACH
CATEGORY OF THE PER

	FRONT OF QUESTIONNAIRE				
ITEM NO.	NOT OBSERVED	UNSATIS-FACTORY	WEAK	ADEQUATE	STRONG
16	1	3	22	114	155
17	3	3	17	90	182
49	2	0	7	86	200
50	0	0	6	112	177
51	8	0	10	137	140
29	3	3	57	149	83
30	17	2	55	181	40
52	35	2	26	157	75
43	22	1	12	139	121
20	28	2	14	128	123
25	16	0	14	113	152
32	102	0	11	109	73
44	52	0	10	139	94
45	68	1	27	131	68
46	136	0	17	108	34
27	17	0	17	142	119
47	144	1	14	92	44
48	16	2	8	149	120
28	21	0	12	138	124
21	133	0	11	86	65
22	107	0	16	102	70
23	94	0	6	98	97
24	22	1	9	138	125
26	45	4	18	119	109
38	63	2	26	138	66
39	99	2	32	123	39
40	25	2	18	168	82
41	105	0	19	110	61
42	74	0	15	134	72
19	135	2	26	89	43
18	101	2	13	118	61
31	133	0	12	109	41
33	70	1	11	141	72
34	200	0	11	61	23
35	142	0	11	104	38
36	99	1	15	115	65
37	18	5	13	137	122

46

ITEM NO.	NOT OBSERVED	BOTTOM 10%	NEXT 40%	NEXT 40%	TOP 10%
		PERSONAL CHARACTERISTICS			
1	0	2	23	93	177
2	1	1	24	100	169
3	0	6	17	84	188
4	0	3	31	112	149
5	0	2	11	76	206
6	1	7	39	122	126
7	6	2	12	100	175
8	0	5	6	94	190
9	1	9	45	122	118
10	0	1	3	79	212
11	25	0	14	104	152
12	0	5	30	111	149
13	0	4	37	100	154
14	0	0	22	126	147
15	9	0	32	143	111

EMPIRICAL DISTRIBUTION HISTOGRAMS

These histograms represent the empirical distribution of score responses for the Performance and Personal Characteristics sections of the PER. Using the cumulative distributions constructed from these histograms, the scores were scaled from 0.0 to 4.0. (Referred to in discussion as the "empirical cumulative distribution transformation").

Performance Scores
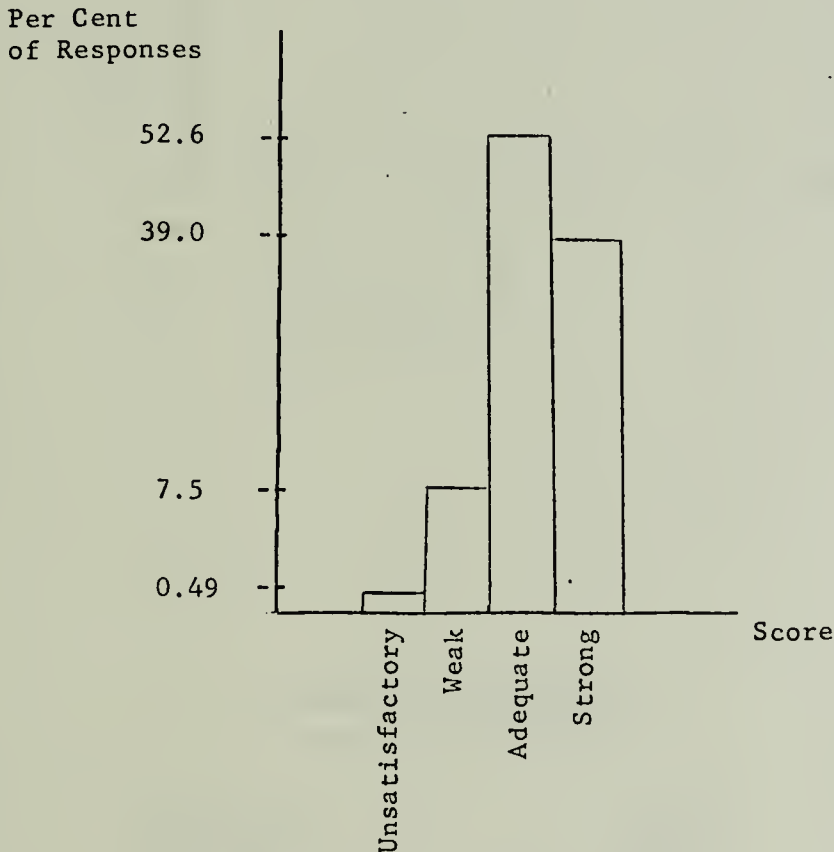
Performance Score Transformation:

Unsatisfactory	.01964
Weak	.31964
Adequate	2.42364
Strong	4.00000

Personal Characteristics Scores
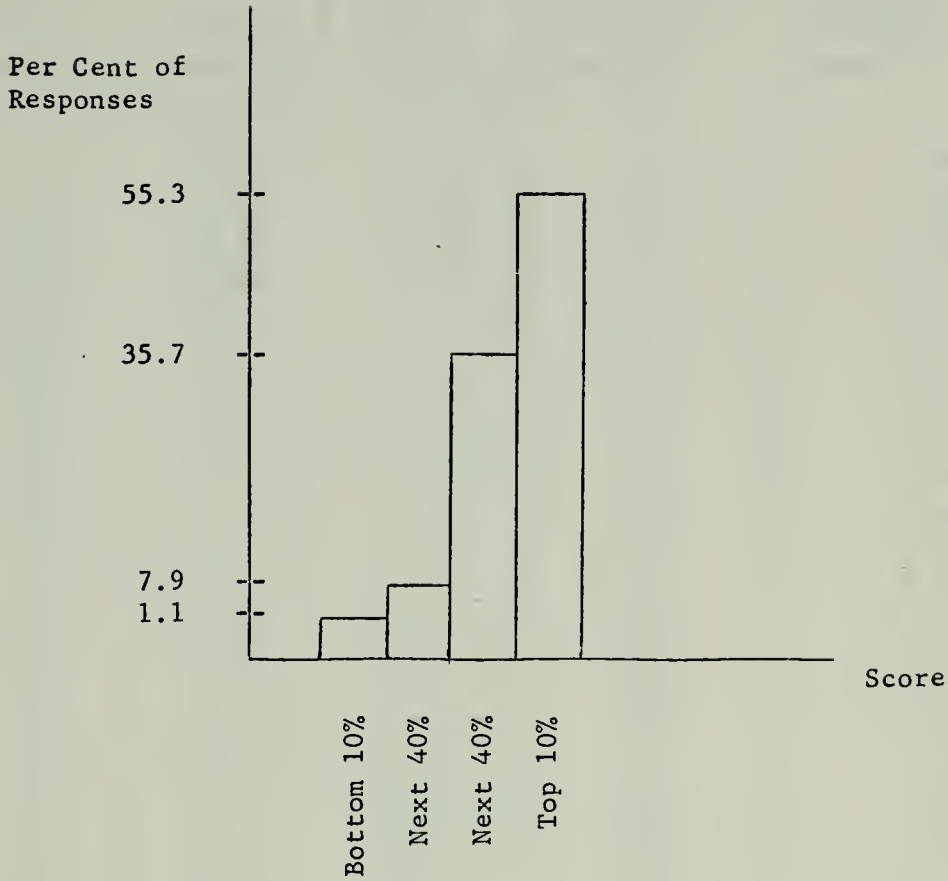
Personal Characteristics
Score Transformation:

Bottom 10%	.044
Next 40%	.360
Next 40%	1.788
Top 10%	4.000

APPENDIX E

AVERAGES OF RANDOM SAMPLE

Subject Number	General Average	Primary Duty Average	Overall Performance Average	Personal Characteristics Average
3	2.62	4.00	2.79	2.56
9	1.90	2.42	1.66	1.08
11	2.88	4.00	3.56	4.00
13	3.80	0.32	2.32	4.00
18	4.00	4.00	4.00	3.26
29	1.63	2.27	1.79	2.29
35	2.80	2.42	2.45	1.96
39	3.21	2.69	2.93	3.12
42	3.61	3.71	3.74	4.00
45	2.82	3.37	2.87	1.94
48	2.35	2.35	2.46	1.99
49	1.63	1.37	1.98	0.88
54	3.41	3.77	3.19	3.71
55	3.32	3.55	3.15	2.32
58	1.52	1.97	2.02	1.03
60	2.55	3.11	2.50	1.31
62	3.61	3.21	3.73	4.00
63	3.41	3.61	3.51	4.00
65	3.61	2.74	3.07	2.67
66	3.61	3.21	3.35	3.85
86	3.01	0.32	2.34	4.00
92	2.82	2.95	2.68	2.67
105	3.55	4.00	3.57	3.37
106	2.16	2.95	2.39	3.71
112	4.00	3.10	3.30	3.85
113	4.00	4.00	3.96	4.00
125	3.41	3.57	3.46	3.26
152	3.41	4.00	3.47	3.26
159	3.55	4.00	3.66	2.73
170	3.61	3.01	3.41	4.00
192	3.21	4.00	3.24	3.26
200	3.10	2.00	2.51	2.14
202	3.80	3.71	3.58	4.00
205	2.09	2.05	1.96	1.12
206	2.55	2.23	2.22	3.41
216	2.09	4.00	2.36	1.55
217	3.34	3.32	3.38	3.56
221	1.90	2.42	2.29	2.53
222	2.82	3.37	3.10	2.04
223	3.80	4.00	3.40	3.56
225	2.87	2.87	2.87	3.56
230	2.09	2.12	2.24	1.61
234	3.80	4.00	3.39	3.12
242	1.63	0.62	1.32	1.73

Subject Number	General Average	Primary Duty Average	Overall Performance Average	Personal Characteristics Average
243	3.01	3.32	3.12	4.00
257	3.21	2.69	3.09	2.67
259	2.42	2.42	2.42	1.79
260	2.49	3.47	2.95	3.12
269	3.21	3.65	3.25	3.71
290	3.01	2.42	2.74	3.12

APPENDIX F

STATISTICS PERFORMED ON PER

H_o: There is no significant difference among the averages of the three sections of the PER (general, primary duty and personal characteristics).

H_1: There is a significant difference among the averages of the three sections of the PER

Let the significant level, α, equal 0.05 and the number of subjects, N, be 50 with k = 3 matched groups.

Since the scores within each of the three matched groups could be ranked, the Friedman two-way analysis of variance was appropriate. Moreover, no normal underlying distribution that would permit the use of the parametric F-test could be assumed.

The following statistic was computed:

$$\chi^2_r = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1)$$

where R_j = sum of ranks for the $j\underline{th}$ group.

Under the null hypothesis, χ^2_r is distributed approximately chi square with k-1 degrees of freedom when N and/or k are large. The region of rejection consists of values of which are greater than 5.99.

The computed value of χ^2_r was 3.01. Therefore, the null hypothesis, H_o was accepted.

H_o: There is no significant difference between the overall performance averages and the personal characteristics averages.

H_1: There is a significant difference between the overall performance averages and the personal characteristics averages.

Let the significance level, α, equal 0.05 and the sample size, N, be 50.

The Wilcoxon Matched-Pairs Signed-Ranks Test was chosen because both the magnitude and direction of the differences between the matched pairs of scores could be determined. Also, no normal underlying distribution that would permit the use of the parametric t-test could be assumed.

The following statistic was computed:

$$z = \frac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)\ (2N+1)}{24}}}$$

where T = sum of the ranks of the differences with the less frequent sign.

Under the null hypothesis, z is distributed as a standard normal statistic. The region of rejection consists of all values of z which are greater in magnitude than 1.96.

The computed z value was -0.2848. Therefore, the null hypothesis, H_o, was accepted.

H_o: There is no significant degree of association among the averages from the three sections of the PER (general, primary duty, and personal characteristics).

H_1: There is a significant degree of association among the averages from the three different sections of the PER.

Let the significant level, α, equal 0.05 and the number of subjects, N, be 50 with k = 3 matched groups of scores.

Since there are three matched groups which can be ranked instead of two, Kendall's coefficient of concordance, W, had to be used. Fortunately, the degree of association as measured by W can be translated into a form comparable to the Spearman rank correlation coefficient. Once again, the assumption of an underlying normal distribution was avoided.

53

The following statistic was computed:

$$\chi^2_w = k\,(N-1)W$$

where $W = \dfrac{12s}{k^2(N^3-N)}$

and s = sum of squares of the observed deviations from the mean of R_j
(sum of the ranks of the $j\underline{th}$ group).

Under the null hypothesis, χ^2_w is distributed approximately chi
square with N-1 degrees of freedom when N is greater than seven. The
region of rejection consists of all values of χ^2_w which are greater
than 70.92.

The computed value of W was 0.6852 yielding a value of χ^2_w equal
to 100.72. Therefore, the null hypothesis, H_o, was rejected.

For purposes of comparison with the next test, a Spearman rank
correlation coefficient equivalent of 0.5278 was computed.

H_o: There is no significant degree of association between the overall
performance averages and the personal characteristics averages.

H_1: There is a significant degree of association between the overall
performance averages and the personal characteristics averages.

Let the significance level, α, equal 0.05 and the number of subjects,
N, be 50.

Since the scores under study could be ranked into two ordered series,
the Spearman rank correlation coefficient, r_s, was chosen to measure the
degree of association between the two groups. Also, no normal underlying
distribution that would permit the use of parametric correlation techniques
could be assumed. Furthermore, there was a desire to compare the degree
of association among the general, primary duty, and personal characteristics
sections of the PER.

The following statistic was computed:

$$t_s = r_s \sqrt{\frac{N-2}{1-r_s^2}}$$

where $r_s = \dfrac{1 - 6 \sum\limits_{i=1}^{N} d_i^2}{N^3 - N}$

and d_i = difference between the matched ranks of subject i.

Under the null hypothesis, t_s is distributed approximately as Student's t with N - 2 degrees of freedom when N is larger than ten. The region of rejection consists of all values of t_s greater than 2.01.

The computed value of r_s was 0.6201 yielding a value of t_s equal to 5.476. Therefore, the null hypothesis, H_o, was rejected.

APPENDIX G

STATICTICS PERFORMED ON THE 16PF AS A PREDICTOR OF PERFORMANCE

Means and Standard Deviations
for Extreme Samples

FACTOR	OVERALL PERFORMANCE ABOVE 3.70		OVERALL PERFORMANCE BELOW 2.00	
	MEAN	STAND. DEV.	MEAN	STAND. DEV.
A	5.49	1.80	5.24	1.69
B	8.09	1.32	7.77	1.95
C	5.32	1.96	5.18	2.16
E	7.09	1.85	7.09	2.05
F	7.28	2.24	7.84	1.57
G	5.53	1.47	4.37	2.02
H	5.47	1.68	5.96	2.13
I	4.73	2.20	5.65	2.19
L	6.17	1.95	6.15	1.76
M	6.18	2.09	6.43	1.65
N	3.37	1.50	3.12	1.41
O	5.85	2.13	5.77	2.74
Q_1	4.53	1.85	5.31	2.07
Q_2	4.46	2.04	4.83	1.72
Q_3	5.97	2.00	5.63	2.68
Q_4	6.80	2.19	6.31	2.60
Q_I	6.80	1.97	7.22	1.90
Q_{II}	6.07	2.08	5.98	2.73
Q_{III}	6.05	1.91	5.64	1.68
Q_{IV}	5.67	1.36	6.41	1.71
Q_V	4.77	2.30	4.83	2.53
Q_{VI}	6.03	1.83	5.70	2.67
Q_{VII}	5.70	1.69	6.25	1.65

FACTOR	PERSONAL CHARACTERISTICS EQUAL TO 4.00		PERSONAL CHARACTERISTICS BELOW 1.56	
	MEAN	STAND. DEV.	MEAN	STAND. DEV.
A	5.16	1.77	4.92	1.65
B	8.02	1.34	8.53	1.37
C	4.95	2.38	5.35	1.89
E	6.95	1.64	7.70	1.81
F	7.22	1.94	7.44	1.89
G	5.17	2.16	3.49	1.61
H	5.44	2.06	5.53	2.14
I	5.44	2.30	5.33	2.40
L	6.26	1.90	6.86	1.56
M	6.35	1.98	6.88	1.73
N	3.03	1.67	3.19	1.45
O	5.92	2.58	5.79	2.32
Q_1	4.96	1.70	5.70	1.95
Q_2	4.50	2.21	5.09	2.18
Q_3	5.70	2.46	5.28	2.30
Q_4	6.66	2.62	6.44	2.07
Q_I	6.69	2.01	6.89	2.19
Q_{II}	6.34	2.73	6.28	2.07
Q_{III}	5.58	1.91	5.90	1.86
Q_{IV}	5.92	1.85	7.30	1.67
Q_V	5.23	2.61	4.72	2.06
Q_{VI}	5.50	2.50	5.32	1.87
Q_{VII}	6.20	1.90	6.90	1.70

The statistical tests presented in this appendix were performed for each of the primary and secondary factors of the 16PF. A table of values of the test statistic for each factor is included.

H_o: There is no significant difference in the distribution of scores between those with personal characteristics averages equal to 4.00 and those with personal characteristics averages below 1.56.

H_1: There is a significant difference in the distribution of scores between those with personal characteristics averages equal to 4.00 and those with unadjusted personal characteristics averages below 1.56.

Let the significance level, α, equal 0.05. The number of subjects with averages equal to 4.00, n_1, equals 49, and the number of subjects with averages below 1.56, n_2, equals 25.

Since two independent samples were compared, the Kolmogorow-Smirnov two-sample test was used to determine whether there was any difference in the distributions from which the two samples were drawn.

The following statistic was computed:

$$D = \max \left| S_{n_1}(x) - S_{n_2}(x) \right| \quad \text{where } S_{n_1}(x) \text{ is the cumulative}$$

distribution function of the ith sample evaluated at x.

The region of rejection consists of all values of D which exceed

$$1.36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 0.394.$$

H_o: There is no significant difference between the subjects whose personal characteristics averages are 4.00 and those whose personal characteristics averages are below 1.56.

H_1: There is a significant difference between the subjects whose personal characteristics averages are 4.00 and those whose personal characteristics averages are below 1.56.

Let the significance level, α, equal 0.05. The number of subjects whose averages are 4.00, n_1, equals 49, where as the number of subjects whose averages are below 1.56, n_2, equals 25.

The Mann-Whitney U Test is one of the most powerful alternatives to the t-test in determining whether two independently chosen samples are drawn from identical populations.

The following statistic was computed:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{(n_1 n_2)(n_1 + n_2 + 1)}{12}}}$$

where $U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$ and R_1 is the sum of the ranks of scores in group 1.

Under the null hypothesis, z is distributed as a standard normal statistic. The region of rejection consists of all values of z which are greater in magnitude than 1.96.

H_0: There is no significant difference between the subjects whose personal characteristics averages are 4.00 and those whose personal characteristics averages are below 1.56.

H_1: There is a significant difference between the subjects whose personal characteristics averages are 4.00 and those whose personal characteristics averages are below 1.56.

Let the significance level, α, equal 0.05. The number of subjects whose averages are 4.00, n_1, equals 49, where as the number of subjects whose averages are below 1.56, n_2, equals 25.

Because the samples might be normally distributed, the parametric t-test was used to test the hypothesis. The region of rejection with

n_1+n_2-2 degrees of freedom consists of all values of t greater than 1.996.

The Kolmogorov-Smirnov two-sample test, the Mann-Whitney U Test, and the t-test were also used to determine significant differences between those whose overall performance averages exceeded 3.70 and those whose overall performance averages were less than 2.00. In all three cases the number of subjects above 3.70, n_1, equalled 22 and the number of subjects below 2.00, n_2, equalled 26. The only other changes to note are in the Kolmogorov-Smirnov Test where the new critical value for D was 0.334 and the t-test where the new critical value for t was 2.0147. The following tables summarize the results of all three tests quite adequately for all factors and both situations.

Kolmogorov-Smirnov Test
(Starred values are significant at the 0.05 level)

Factor	Overall Performance D	Personal Characteristics D
A	0.147	0.207
B	0.182	0.256
C	0.199	0.209
E	0.070	0.234
F	0.241	0.129
G	0.395*	0.433*
H	0.249	0.059
I	0.255	0.109
L	0.077	0.255
M	0.105	0.192
N	0.160	0.171
O	0.178	0.108
Q_1	0.199	0.276
Q_2	0.192	0.232
Q_3	0.255	0.193
Q_4	0.178	0.127
Q_I	0.249	0.108
Q_{II}	0.263	0.110
Q_{III}	0.178	0.124
Q_{IV}	0.196	0.293
Q_V	0.122	0.189
Q_{VI}	0.172	0.127
Q_{VII}	0.203	0.313

Mann-Whitney U Test
(Starred values are significant at the 0.05 level)

Factor	Overall Performance z	Personal Characteristics z
A	-0.363	-0.624
B	-0.272	-1.400
C	-0.352	-1.010
E	-0.042	-1.980*
F	-0.825	-0.396
G	-2.285*	-3.372*
H	-0.787	-0.217
I	-1.483	-0.200
L	-0.125	-1.347
M	-0.425	-1.214
N	-0.385	-0.639
O	-0.073	-0.149
Q_1	-1.359	-1.659
Q_2	-1.109	-1.145
Q_3	-0.187	-0.911
Q_4	-0.633	-0.664
Q_I	-0.735	-0.606
Q_{II}	-0.052	-0.509
Q_{III}	-0.787	-0.560
Q_{IV}	-1.749	-3.030*
Q_V	-1.149	-0.949
Q_{VI}	-0.362	-0.160
Q_{VII}	-1.087	-1.835

t-Test
(Starred values are significant at the 0.05 level)

Factor	Overall Performance t	Personal Characteristics t
A	0.496	0.564
B	0.653	-1.537
C	0.233	-0.730
E	0.000	-1.796
F	-1.015	-0.465
G	2.237*	3.429*
H	-0.873	-0.176
I	-1.447	0.192
L	0.037	-1.361
M	0.463	-1.135
N	0.594	-0.407
O	0.111	0.212
Q_1	-1.365	-1.685
Q_2	-0.682	-1.169
Q_3	0.491	0.710
Q_4	0.699	0.365
Q_I	-0.750	-0.393
Q_{II}	0.127	0.097
Q_{III}	0.791	-0.677
Q_{IV}	-1.638	-3.132*
Q_V	-0.085	0.850
Q_{VI}	0.490	0.317
Q_{VII}	-1.138	-1.551

BIBLIOGRAPHY

Anastasi, A., Psychological Testing, Macmillan Company, 1954.

Buros, O. K., Personality Tests and Reviews, Gryphon Press, 1970.

Buros, O. K., The Sixth Mental Measurements Yearbook, Gryphon Press, 1965.

Cattell, R. D., The Scientific Analysis of Personality, Penguin Books Ltd., 1965.

Cattell, R. B., Eber, H. W., and Tatsuoka, M. M., Handbook for the 16PF, Institute for Personality and Ability Testing, 1970.

Cronbach, L. J., Essentials of Psychological Testing, Harpor and Bros., 1960.

Freeman, F. S., Theory and Practice of Psychological Testing, Henry Holt and Company, 1955.

Kirk, R. E., Experimental Design - Procedures for the Behavioral Sciences, Wadsworth Publishing Company, 1968.

Lauzon, G. P., A Computer - Graphics Separation Algorithm for Pattern Classification and Cluster Analysis, Masters Thesis, Naval Postgraduate School, 1973.

Murstein, B. I., Handbook of Projective Techniques, Basic Books Inc., 1965.

Nie, N. H., Bent, D. H., and Hull, C. H., SPSS: Statistical Package for the Social Sciences, McGraw-Hill Book Co., 1970.

Nunnally, J. C., Psychometric Theory, McGraw-Hill Book Company, 1967.

Nunnally, J. C., Educational Measurement and Evaluation, McGraw-Hill Book Company, 1972.

Oppenheim, A. N., Questionnaire Design and Attitude Measurement, Basic Books, Inc., 1966.

Pullen, J. R., A Comparative Study of Personality Factors and Certain Other Variables, University Microfilms, 1971.

Siegel, S., Non-Parametric Statistics, McGraw-Hill Book Company, 1956.

Vernon, P. E., Personality Assessment: A Critical Survey, John Wiley and Sons Inc., 1964.

INITIAL DISTRIBUTION LIST

No. Copies

1. Library, Code 0212   2
    Naval Postgraduate School
    Monterey, California   93940

2. Asst. Professor D. E. Neil (thesis advisor)   1
    Department of Operations Research and
    Administrative Sciences
    Naval Postgraduate School
    Monterey, California   93940

3. Ensign Jeffrey Wiley Crews, USN (student)   1
    1260 N. E. 8th Street
    Ocala, Florida   32670

4. Ensign Jeffrey Willian Cummings, USN (student)   1
    USS Goldsborough DDG-20
    FPO San Francisco, California   96601

5. Dr. Karel Montor
    Department of Naval Command and Management   1
    U. S. Naval Academy
    Annapolis, Maryland   21402

6. LCDR J. J. Reed   1
    Operations Analysis Study Group
    U. S. Naval Academy
    Annapolis, Maryland   21402

7. Chief of Naval Personnel (Pers-11b)   1
    Department of the Navy
    Washington, D. C.   20370

8. Department of Operations Research and   1
    Administrative Sciences
    Code 55
    Naval Postgraduate School
    Monterey, California   93940

9. Assoc. Professor J. K. Arima (second reader)   1
    Department of Operations Research and
    Administrative Sciences
    Naval Postgraduate School
    Monterey, California   93940

10. Defense Documentation Center   2
    Cameron Station
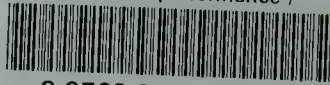   Alexandria, Virginia   22314